

Petrus Hispanus Lectures 2014

Why Theories of Concepts Should Not Ignore the Problem of Acquisition

Susan Carey
Harvard University¹

BIBLID [0873-626X (2015) 41; pp. 113-163]

Abstract

A theory of conceptual development must provide an account of the innate representational repertoire, must characterize how these initial representations differ from the adult state, and must provide an account of the processes that transform the initial into mature representations. In *The Origin of Concepts* (Carey 2009), I defend three theses: (1) the initial state includes rich conceptual representations, (2) nonetheless, there are radical discontinuities between early and later developing conceptual systems, (3) Quinean bootstrapping is one learning mechanism that underlies the creation of new representational resources, enabling such discontinuity. Here I argue that the theory of conceptual development developed in *The Origin of Concepts* constrains our theories of concepts themselves, and addresses two of Fodor's challenges to cognitive science; namely, to show how learning could possibly lead to an increase in expressive power and to defeat Mad Dog Nativism, the thesis that all concepts lexicalized as mono-morphemic words are innate. In response to Fodor, I show that, and how, new primitives in a language of thought can be learned, that there are easy routes and hard ones to doing so, and that characterizing the learning mechanisms in each illuminates how conceptual role partially determines conceptual content.

Keywords

Concept, concept acquisition.

¹ Reprinted courtesy of The MIT Press from *The Conceptual Mind*, edited by Eric Margolis and Stephen Laurence, Cambridge, MA: MIT Press, 2015 (<https://mitpress.mit.edu/index.php?q=node/249683>).

The *Petrus Hispanus Lectures 2014* were delivered by Professor Susan Carey at the University of Lisbon on May 27th and 29th 2014.

Disputatio, Vol. VII, No. 41, November 2015

Received: 05/11/2015

1 Introduction

The human conceptual repertoire is a unique phenomenon on earth, posing a formidable challenge to the disciplines of cognitive science. Alone among animals, humans can ponder the causes and cures of pancreatic cancer and global warming. How are we to account for the human capacity to create concepts such as CLIMATE, CANCER, ELECTRON, INFINITY, GALAXY and WISDOM? How do such concepts arise, both over history and in ontogenesis? Rightly, most attempts to provide such an account center on what makes concept attainment possible, but the literature on concept development adds a second question. Why is concept attainment (sometimes) so easy and what (sometimes) makes concept attainment so hard? Easy: some new concepts are formed upon first encountering a novel entity or hearing a new word in context (Carey 1978). Hard: others emerge only upon years of exposure, often involving concentrated study under metaconceptual control, and are not achieved by many humans in spite of years of explicit tutoring in school (Carey 2009). Considering what underlies this difference illuminates both how concepts are attained and what concepts are.

A theory of conceptual development must have three components. First it must characterize the innate conceptual repertoire—the representations that are the input into subsequent learning processes. Second, it must describe how the initial stock of representations differs from the adult conceptual system. Third, it must characterize the mechanisms that achieve the transformation of the initial into the final state.

The two projects of constructing a theory of concept acquisition and constructing a theory of concepts fit within a single intellectual enterprise. Obviously, a theory of concept acquisition must be consistent with what concepts *are*. But the relation between the two projects goes both ways, a fact that has played almost no role in the psychological literature on concepts (see, for example, the excellent reviews in Smith and Medin 1981, and in Murphy 2002). With the exception of developmental psychologists, cognitive scientists working on concepts have mostly abandoned the problem of characterizing and accounting for the features that enter into their learning models, often coding them with dummy variables.

This was not always so. For example, in theorizing about concepts, the British Empiricists made accounting for acquisition a central concern. They, like many modern thinkers, assumed that all concept learning begins with a primitive sensory or perceptual vocabulary. That project is doomed by the simple fact that it is impossible to express most concepts in terms of perceptual features (e.g., CAUSE, GOOD, SEVEN, GOLD, DOG...). In response, some theorists posit a rich stock of innate conceptual primitives, assuming that the adult conceptual repertoire can be built from them by conceptual combination. That is, they assume that the computational primitives that structure the adult conceptual repertoire and the innate primitives over which hypothesis testing is carried out early in development are one and the same set (e.g., Levin and Pinker 1991; Miller 1977; Miller and Johnson-Laird 1976). A moment's reflection shows this assumption is also wrong. For example, the definition of GOLD within modern chemistry might be ELEMENT WITH ATOMIC NUMBER 79. Clearly the theoretical primitives ELEMENT and ATOM are not innate conceptual features, as they arise in modern chemistry and physics only in the 18th and 19th centuries, after many episodes of conceptual change. (Of course, it is an open question whether ELEMENT and ATOM are definable in terms of developmental primitives; there are no proposals for possible definitions in terms of innately available primitives). Or take the features that determine the prototype structure of animal concepts (e.g., BIRD: FLIES, LAYS EGGS, HAS WINGS, NESTS IN TREES, HAS A BEAK, SINGS,...). Participants in studies provide just these when asked to list the features of birds. Furthermore, overlap in these features with others at this grain predicts judged similarity of birds to other animals, and overlap in particular values of them (e.g., beak type), as well as other features such as color and size, predicts prototypicality within the category of birds. That is, this feature space definitely underlies adult prototypicality structure. Prototype learning models assume that learning a new concept involves constructing a summary representation of a category in terms of such features, and then using this summary representation to probabilistically determine category membership. But a moment's reflection shows these models just help themselves to features that are not, for the most part, innate primitives—many are no less abstract nor no less theory-laden than the concept BIRD itself.

In a recent book (Carey, 2009, *The Origin of Concepts*, hereafter, *TOOC*), I take on the dual projects of accounting for conceptual development and characterizing the nature of human concepts. Towards a theory of conceptual development, I defend three theses. With respect to the initial state, contrary to historically important thinkers such as the British empiricists, Quine, and Piaget, as well as many contemporary scientists, the innate stock of primitives is not limited to sensory, perceptual or sensory-motor representations. Rather, there are also innate conceptual representations, embedded in systems of core cognition, with contents such as AGENT, OBJECT, GOAL, CAUSE, and APPROXIMATELY 10. With respect to developmental change, contrary to continuity theorists such as Fodor (1975), Pinker (2007) and many others, there are major discontinuities over the course of conceptual development. By ‘discontinuity’ I mean qualitative changes in representational structure, in which the later emerging system of representation cannot be expressed in terms of the conceptual resources available at the earlier time. Conceptual development consists of episodes of qualitative change, resulting in systems of representation with more expressive power than, and sometime incommensurable with, those from which they are built. Increases in expressive power and incommensurabilities are two types of conceptual discontinuities. With respect to a learning mechanism that achieves conceptual discontinuity, I offer Quinian bootstrapping.

Toward a theory of concepts that meshes with the picture of conceptual development in *TOOC*, I support dual factor theory (e.g., Block 1986). The two factors are sometimes called ‘wide’ and ‘narrow’ content. The wide content of our mental representations is partly determined by causal connections between mental symbols, on the one hand, and the entities to which they refer. To the extent this is so, all current psychological theories of concepts are on the wrong track—concepts are not prototypes, exemplar representations, nor theories of the entities they represent. However, contrary to philosophical views that deny that meanings are determined in any way by what’s in the head (e.g., Dretske 1981, Fodor 1998, Kripke 1972/1980, Putnam 1975), *TOOC* argues that some aspects of inferential role are content determining (narrow content). The challenge for psychologists is saying what aspects of mental representation of

entities we can think about partly determine the meaning of concepts of those entities, and which are simply what we believe about those entities (sometimes called the project of distinguishing concepts from conceptions, Rey 1983). Facts about conceptual development constrain a theory of narrow content.

While the goal of *TOOC* was to explicate and defend the above three theses about conceptual development and sketch how they mesh with a dual factor theory of concepts, I also addressed Fodor's (1975, 1980) two related challenges to cognitive science—first, to show how learning can possibly result in increased expressive power, and to defeat the conclusion that all concepts lexicalized as monomorphemic words are innate. The key to answering both of these challenges, as well as to understanding conceptual discontinuities in general, is to show that, and how, new conceptual *primitives* can be learned. Conceptual primitives are the building blocks of thought, the bottom level of decomposition into terms that articulate mental propositions and otherwise enter into inference. Conceived of this way, there is no logical requirement that conceptual primitives cannot be learned.

Rey (2014) denies that the project is successful in meeting Fodor's challenges, as do Fodor (2010) and Rips and colleagues (Rips et al. 2008, 2013). Although I ultimately disagree, I appreciate many of the points these critics make along the way. These debates bring into focus how the projects of understanding conceptual development and understanding the nature of concepts, learning, and the human mind are intertwined. In this paper I lay out these debates on the interrelated issues of conceptual discontinuity, increases in expressive power, and Quinian bootstrapping and begin to sketch how they bear on our understanding of the nature of concepts. I show how new primitives can be learned, and how this fact bears on these debates.

2 The dialectic according to Fodor, Rey and Rips et al.

A kind of logical constructivism is at the heart of Fodor's and Rey's (and at least implicitly) Rips et al.'s dialectic. These writers, like many others, take expressive power to be a function of innate primitives, and what can—in principle if not in fact—be built from them

using the resources of the logic available to the learner. Expressive power is a logical/semantic notion. So long as the characterization of learning mechanisms is exhausted by specifying the set of innate primitives and the logical resources through which one builds new representations from those primitives, clearly one cannot increase expressive power by learning (Fodor 1980).

My response to this picture of learning and conceptual development is to argue that learning mechanisms can create new primitives, new primitives that cannot be constructed from antecedently existent primitives by logical combination, and thus increase the expressive power of the conceptual system. In addition, my concern is with how new primitives actually come into being; if there are processes that yield new primitives, then the question is whether such processes actually underlie the emergence of any given representation.

Fodor's (1975) second challenge to cognitive science is to defeat his argument for Mad Dog Nativism, that is, to defeat the argument that virtually all of the over 500,000 concepts lexicalized by mono-morphemic words in the Oxford English Dictionary are innate. Rey (2014) lays out Fodor's argument as follows:

Premise 1: (Hypothesis Confirmation). All learning is hypothesis confirmation.

Premise 2: (Logical Construction) One can learn new concepts only by creating and confirming hypotheses formulated in terms of logical constructions from antecedently available primitive concepts.

Premise 3: (Atomism). The concepts underlying mono-morphemic words cannot be analyzed as logical constructions of other concepts, primitive or otherwise. (Actually, Fodor says 'most' mono-morphemic concepts cannot be so analyzed, but for simplicity I will assume 'all' rather than 'most').

Conclusion: (Innateness). In order to acquire a new concept lexicalized as a mono-morphemic word, one would have to confirm hypotheses already containing the concept to be learned. Therefore, no such concept can be learned.

TOOC answers this challenge by giving reasons to deny premises 1 and 2. My basic strategy has been to provide several case studies of transitions between conceptual systems in which the later one expresses concepts that are not logical constructions from the earlier one (Carey 1985, 2009; Smith, Carey and Wiser 1985; Wiser and Carey 1983). Sometimes this is because of local incommensurability, as in case studies of thermal concepts, biological concepts and electromagnetic concepts in the history of science, or concepts of matter/weight/and density in intuitive physics in childhood and the concepts of life and death in childhood). Sometimes it is because of developments within mathematic representations that increase expressive power without necessarily involving local incommensurability (as in case studies of the origins of concepts of integers and rational number).² *TOOC* then goes on to analyze how Quinian bootstrapping plays a role in transitions of both types.

The central issue dividing my views from the critics I focus on here is discontinuity. These critics deny the very possibility of conceptual discontinuities, as well as offering a positive view of conceptual development in terms of Premises 1 and 2 of Fodor's argument which they claim shows how conceptual development is possible without discontinuity. Rips and his colleagues suggest that claims for discontinuities are incompatible with claims that concepts are learned (Rips and Hespos 2011; Rips, Asmuth and Bloomfield 2013). Again, the key is understanding that, and how, new conceptual primitives can be learned. These critics argue that my proposal for a learning mechanism that can underlie conceptual discontinuity, Quinian bootstrapping, fails, partly through failing to confront a psychologized version of Goodman's new riddle of induction (Rey 2014, Rips et al. 2008).

With respect to Rips' and his colleagues worries that concept learning and concept discontinuity are incompatible, let me clarify what the debate is *not* about. The existence of conceptual discontinuity cannot entail that it is impossible for an organism to acquire some

² The case study of the construction of the integers is the focus of Rey's, Rips et al.'s, and Fodor's critiques. I will discuss whether this episode of conceptual development truly involves a discontinuity, and an increase of expressive power, when I turn to it in Sections 8 and 9 below.

later representations, given its initial state, except through maturation or magical processes that don't involve learning (e.g., being hit on the head). What is actual is possible. The mechanisms (there are many) that underlie the acquisition of our representational repertoire, in general, and our conceptual repertoire in particular, if they are learning mechanisms, are computational processes. At stake are premises 1 and 2 of Fodor's argument, which all of these critics explicitly or implicitly endorse. I agree that most of conceptual development consists of hypothesis confirmation, where the hypotheses are articulated in terms of already available concepts. Discontinuities arise in episodes of conceptual development where this is not the right model.

With respect to the positive proposal, Mad Dog Nativism requires that virtually all the 500,000 concepts lexicalized in English, plus those that will come to be lexicalized in the future, are innate, existing in some way in the infant's mind. This isn't comforting as a positive proposal that obviates the need for concept learning. *A priori*, it is highly unlikely that QUARK and CARBURETOR and FAX are innate concepts, existing in some kind of hypothesis space available for hypothesis testing. Noting this unlikelihood, Rey (2014) distinguishes between manifest concepts (those currently available for hypothesis testing and inference) and what he calls 'possessed' concepts (those that exist in the mind in some way, but are not currently available for thought, or those that can be constructed, by logical combination from that initial set). Rey defines possessed concepts as those that have the *potential* to be manifest. Here I use 'potential' concepts instead of 'possessed' concepts to express this notion. Nobody would ever deny that an actual manifest concept had the potential to be the output of some developmental process, and in the light of characterizations of those developmental processes, we can and do explore the representational repertoire it can achieve. Exploring the possible outputs of the learning mechanisms we investigate is an important part of characterizing these mechanisms. Calling the potential output of concept learning mechanisms 'possessed concepts' implies something stronger, that they exist somehow in the mind prior to becoming manifest. Of course, Premises 1 and 2 specify one way we can think about this stronger notion 'possession:' the innate primitives, along with the combinatorial apparatus of logic and language

constitutes a space of alternative hypotheses about which concepts apply in particular contexts (e.g., to support the meaning of a word), and this space exhausts the potential concepts that are attainable. The writers I am criticizing here assume that potential concepts constitute a space of alternatives, laying in wait to become manifest, and that manifestation consists in *being* or *being logically constructed* from these innately possessed primitives. These assumptions follow from premises 1 and 2 of Fodor's argument, the premises I deny.

3 Initial response

My project concerns manifest concepts. To reiterate, manifest concepts are those currently available to for thought, inference, and guiding action. The developmental primitives I study are those we can find evidence for in the baby's or animal's behavior. They must be available to support inference and action in order to be diagnosed, i.e., they must be manifest (currently available for thought). In what follows I argue that concept manifestation is where the debates about expressive power, conceptual continuity/discontinuity, and induction *actually* play out.

For any representational system we posit, we are committed to there being answers to three questions. First, what is the format of the symbols in the system; second, what determines their referents; and third, what is their computational role in thought. A worked example in *TOOC* is the evolutionarily ancient system of number representations in which the mental symbols are quantities (rates of firing, or size of populations of neurons) that are linear or logarithmic functions of the cardinal values of sets, which in turn are input into numerical computations such as number comparison, addition, subtraction, multiplication, division, ratio calculations, probability calculations, and others (see Dehaene 1997, for a book-length treatment of this system of numerical representations). We can only explore such systems with psychological methods that diagnose manifest representations. The project of *TOOC* is understanding the representational resources available as the child or adult interacts with the world, how these arise and change over development. These representations are the ones available for hypothesis testing, as input into further learning, and to play a computational role in thought.

And it is successive manifest conceptual systems one must analyze to establish qualitative changes (i.e., conceptual discontinuities).

In what follows I flesh out these points, explicating how *TOOC* attempts to answer Fodor's challenges to cognitive science. The issues include a characterization of the nature of learning (Fodor's first premise), the unjustified acceptance of the logical construction model as the only model of concept learning (Fodor's second premise), the misleading analogy of the totality of concepts ultimately attainable as a hypothesis space, the characterization of how primitives arise (both in cases where this is easy and in cases where this is hard), and the characterization of constraints of induction (and constraints on learning more generally, in cases where learning does not involve induction).

Let me begin with the premises in Fodor's argument that I deny. I first comment on why these premises matter and I then show why they are wrong.

4 Premise 2. Logical construction

The premise that all concepts must either be innate or buildable by combination from innate primitives through innate logical combinatorial devices is widely adopted within cognitive science. For example, the dominant theoretical project within the field of lexical development in the 1970s was to attempt to discover the lexical primitives in terms of which lexical items are defined, and to study the intermediate hypotheses children entertain as they construct new concepts from those primitives (see Carey 1982, for a review and critique). That is, it was just assumed that definitional primitives are innate. There I called this view 'piece by piece construction'; Margolis and Laurence (2011) call it 'the building blocks model'. Here, I will call it 'the logical construction model', in honor of Premise 2. In contrast, I argue (Carey 1982, *TOOC*) that computational primitives need not be innate. They can be acquired through learning processes that do not consist of logical construction from innate primitives.

One central issue is atomism. If many of the primitives in adult thought (e.g., the concepts expressed by words like 'dog' or 'cancer'), cannot be defined in terms of innately manifest concepts, then

they either must be innate primitives or it must be possible to learn computational primitives through some mechanism that does not consist of building new concepts by logical combination of antecedently available ones, and is not exhausted by confirming a hypothesis stated in terms of the to be acquired concept. I accept Fodor's arguments that most lexical concepts are definitional primitives.

Notice that the possibility one can learn new primitives matters to the question of expressive power of the system. The expressive power of a system of representations is a function of its atomic terms and combinatorial apparatus. The logical connectives and operators (sentential operators, modals, quantifiers) are not the only primitives that matter to expressive power. If DOG cannot be logically constructed from primitives, then acquiring the concept DOG increases expressive power of the system (see Weiskopf 2008). That is, non-logical primitives figure into semantic/logical expressive possibilities as well as do logical ones. This is one reason that the question of whether one *learns* the concept DOG is so central to the debate between Fodor and his critics.

5 Premise 1. All learning is hypothesis formulation and testing

To evaluate this proposition we must agree upon what hypothesis testing is and what learning is. Bayesian models specify the essence of hypothesis testing algorithms. Hypothesis testing requires a space of antecedently manifest concepts, each associated with prior probabilities, and each specifying likelihood functions from any possible evidence to the probability that it supports any given hypothesis. Hypothesis testing then involves choosing among the alternative hypotheses on the basis of evidence. Fodor (1975, 2008) claims that all learning mechanisms reduce to hypothesis testing, at least implicitly. I agree that any learning mechanism that revises representations as evidence accumulates (e.g., associative mechanisms that update strengths of association, supervised learning algorithms such as connectionist back propagation) do indeed do so. However, as Margolis and Laurence (2011) point out in a reply to Fodor's 2008 book (*LOT2*), a cursory examination of the variety of attested learning mechanisms in the animal kingdom shows that the generalization that *all* learning mechanisms reduce to hypothesis confirmation is

wildly off the mark. Rote learning (memorizing a phone number), one-trial associational learning (e.g., the Garcia effect, the creation of a food aversion as a result of becoming nauseous some fixed time after having eaten a novel food, Garcia et al. 1955), and many other types of learning do not involve choosing among multiple hypotheses, confirming one of them, in the light of accumulating evidence. And as we shall see, such mechanisms have roles to play in creating new conceptual primitives.

Of course, the claim that these are learning mechanisms depends upon what one takes learning to be. Learning mechanisms share a few essential properties that allow us to recognize clear examples when we encounter them. All learning results in representational changes in response to representational inputs, where those inputs can be seen (by the scientist) to provide evidence relevant to the representational change. That is, learning is a computational process, requiring representational inputs that can be conceptualized as providing relevant *information*. Sometimes, as in the case of explicit or implicit hypothesis testing, the organism itself evaluates the information in the input with respect to its evidential status (as in all forms of Bayesian learning mechanisms). But other times, the learning mechanism is a domain specific adaptation that responds to information by simply effecting a representational change of relevance to the organism—an example being the learning mechanism that underlies the Garcia effect mentioned above. No further evidence is evaluated, so there is no hypothesis confirmation.

6 The relatively easy route to new representational primitives: domain specific learning mechanisms

The problem of acquisition arises in the case of any representation, conceptual or otherwise, that end up in the manifest repertoire of an animal. The literatures of psychology and ethology have described hundreds of domain-specific learning mechanisms that simply compute new representations from input, having arisen in the course of natural selection to do just that. Most of these representations are not conceptual ones, but considering how they are acquired shows that the learning mechanisms involved do not always involve hypoth-

esis testing, thus providing counterexamples to Premise 1. They also do not implement logical construction from primitives, and thus provide counterexamples to Premise 2. Considering how they work illuminates why it's a mistake to consider potential representations as a space of existent representations, ready to be *chosen among* or *built from* in a process of manifestation.

TOOC's example of an evolved domain-specific learning mechanism is that which underlies Indigo buntings' *learning* which part of the night sky indicates north. This matters crucially to Indigo buntings, for they migrate over 3500 miles each spring (north) and fall (south), and they navigate by the stars. Because the earth tilts back and forth on its axis, what part of the night sky indicates north changes radically on a 30,000 year cycle. Sometime not too far in the future, the north star will be Vega, not Polaris. Thus, it is unlikely that an innate representation of Polaris as the north star was created by natural selection, and indeed, Steven Emlen (1975) discovered the learning mechanism through which Indigo buntings create the representation of north that will play such a crucial role in their migratory life. The learning device that achieves this analyzes the center of rotation of the night sky, and stores the configuration of stars that can allow the bird to recognize the position of north from a static sighting (as it has to do every time it starts to fly during its migrations in the spring and the fall, and as it monitors its course).

This mechanism computes what it is designed to compute—nothing more, nothing less. It creates an essential representation in the computational machinery of Indigo buntings, the specification of north in the night sky. Of course, there is a prepared computational role for this representation, but the representation of north as specified by the stars must still be learned, and is an essential primitive in the computational machinery underlying Bunting navigation. Domain specific learning mechanisms of this sort are often supported by dedicated neural machinery that atrophies after its work is done, leading to critical periods. This is such a case; if a bird is prevented from seeing the night sky as a nestling, no amount of exposure to the rotating night sky later in life allows the bird to identify north, and the bird perishes.

This example is worth dwelling upon with respect to whether representations that can be achieved should be thought of as part of

an existing space of hypotheses, and whether the acquisition mechanism involves hypothesis confirmation or logical combination. Until the learning episode is completed, there is no manifest representation that specifies north in the night sky in the bird's mind. However, this learning mechanism can learn any of a very large number of star configurations constellations that could indicate north. Indeed, part of the evidence that this *is* the learning mechanism through which indigo buntings establish Polaris as the north star are planetarium experiments in which the night sky is made to rotate around an arbitrarily chosen part of the night sky while the birds are nestlings. The birds then use the north star so specified to set their course when it's time to migrate. Thus, there are a plethora of potential north stars. And clearly, one can investigate limits on the system (e.g., if stars were equally distributed throughout the sky, or if they were too densely packed to be resolved, or if the patterns of stars showed large scale repetitions, this couldn't work.) It is only with an actual representational/computational characterization of this learning mechanism that the space of potential north stars the Bunting could acquire representations of can be explored. Such is always the case.

What about hypothesis testing? I take the essential features of hypothesis testing to be two: (1) the learning mechanism must entertain alternatives, and (2) choice among them must be based on evidence. The space of potential representations of north that can be achieved by Buntings is in *no* way a hypothesis space. In no way does an Indigo Bunting's acquiring a representation of north consist of choosing among possibilities. Calling the possible specifications of north a 'hypothesis space' is wildly misleading. There is no initial set of possibilities, with associated priors, with likelihood functions associated with them. The animal never considers any possibility other than the output of the learning mechanism, and the animal has no way of testing whether the specification of north that is the output of the learning mechanism is actually NORTH. The bird simply computes it, and lives or dies by it.

This case is also worth dwelling upon with respect to the other issues on the table. Not only does this case not involve hypothesis formulation and testing, it also does not involve building a new representation out of primitives by logical combination. And since there is no induction involved, the issues of constraints on induction do not

arise. Of course, all learning mechanisms must be highly constrained to be effective, and characterizing real learning mechanisms allows us to understand the constraints under which they operate. This is a highly constrained learning mechanism; it considers only one kind of information to create a representation that has only one computational role. It is of no use to the bird in helping the bird learn what to eat, who to mate with, or where its nest is in a local environment.

Navigation is not a special case. There have been hundreds of such domain specific learning mechanisms detailed in the literatures of ethology and psychology, including the imprinting mechanisms that allow infants (animals and humans) to identify conspecifics in general and their caretakers in particular, mechanisms that allow animals to learn what food to eat (the Garcia effect just one of dozens of domain specific learning mechanisms through which omnivores like rats and humans achieve this feat), bird song learning, and so on (see Gallistel et al. 1991, for a review of four such domain-specific information-expectant learning mechanisms, and Gallistel 1990 for a nuanced discussion of the nature of learning).

In sum, the animal literature provides many examples of learning mechanisms designed to form new computational primitives, learning mechanisms that implicate neither logical construction from existing primitives (Premise 2), nor hypothesis testing and confirmation (Premise 1). One can (and one does) explore the space of possible outputs of these mechanisms, for this is one way they can be fully characterized and their existence empirically tested, but in no way is there a space of representations laying in wait, existing ready to be manifested, existing ready to be chosen among.

7 The relatively easy route to new conceptual primitives

The learning mechanism described above acquires a new primitive representation, a representation that allows the animal to identify north in the night sky, to guide navigation. One might argue it is not a new *conceptual* representation. Its format is surely iconic, and its computational role is both highly domain specific and sensori-motor. There are, however, learning mechanisms that similarly respond to inputs of certain types by simply creating new conceptual primitives, primitives that enter into representations with propositional format

and participate in the full productivity of language and causal inference. These domain specific concept learning mechanisms need not involve hypothesis testing, and do not involve constructing new concepts by logical combination. Take the Block (1986)/Macnamara (1986)/Margolis (1998) object-kind learning mechanism for example.³ This learning mechanism is triggered by encountering a novel object (as specified by core cognition of objects) with obviously non-arbitrary structure. As Prasada et al. (2002) showed, there are several cues to non-arbitrary structure: the object has complex yet regular shape (e.g., symmetries, repetition), or there are multiple objects that share a complex irregular shape, or the object has functionally relevant parts, or the object recognizably falls under an already represented superordinate kind (e.g., kind of agent, kind of animal, kind of artifact). Core cognition contains perceptual input analyzers that are sensitive to cues to each of these properties of individual objects. Encountering an individual with one or more of these properties triggers establishing a new representational primitive that can be glossed SAME BASIC LEVEL KIND AS THAT OBJECT. Reference to the kind is ensured by representation of the surface properties of the individual or individuals that occasioned the new concept (and these represented surface properties get enriched and even overturned as bases of reference and categorization as more is learned about the kind). The content of the new concept depends upon the referent, the conceptual role provided by the basic level kind schema (psychological essentialism), and the conceptual roles provided by any superordinate kind schemas that the individual is taken to fall under (e.g., AGENT, ANIMAL, ARTIFACT, these in turn being constrained by their roles in different systems of core cognition or constructed theories).

Consider encountering a kangaroo for the first time. Such an encounter might lead to the formation of a concept KANGAROO that

³ These writers discuss this mechanism as a *natural kind* learning mechanism (e.g., kinds of animals or kinds of plants), but I believe the domain of this mechanism is object kind representations (as opposed to object properties, individual objects, or the events in which objects participate). Roughly, kind representations are inductively deep, and kinds are construed in accordance with the constraints that constitute psychological essentialism in Strevens' (2000) sense. Artifact kinds fall under the domain of this mechanism as well as do natural kinds (Kelemen and Carey 2007).

represents animals that are the same basic level kind as the newly encountered one. No enumerative induction is needed; the concept is what Strevens (2009) calls ‘introjected’ into one’s set of primitives. This concept, falling under psychological essentialism (as it is a kind concept), reflects the many constraints on kind concepts. That is, the conceptual role SAME KIND AS includes assumptions that something causes the non-random structure that triggered the formation of the new concept, that these underlying causes are shared by all members of the kind (now, in the past, in the future), that the surface properties that specify the individual that occasioned the new concept may not hold for all members, possibly not even typical members. Furthermore, the current guesses about the nature of the relevant causal mechanisms relevant to the creation of members of this kind, to determining their properties, and to tracing numerical identity though time, are taken to be open to revision. That is, there is no definition that determines membership in the kind; learners treat everything they represent about the kind up for revision (including, even that there IS a new kind—the individual we encountered might have been a mutant raccoon).

This mechanism creates new primitives, not definable in terms of other manifest concepts, and thus increases the expressive power of the conceptual system. The concept KANGAROO is not definable in terms of antecedently available primitives using the combinatorial machinery of logic. Before creating this concept, one could not think thoughts about kangaroos, just as before analyzing the center of rotation of the night sky and storing a representation of NORTH so specified, an Indigo bunting could not set or guide a course of flight toward or away from north. Of course the kind learning mechanism ensures that creating new primitives for kinds is easy; one need only encounter an individual that one takes to be an individual of a new kind, and store a representation of what that individual looks like. But this process involves neither induction nor hypothesis testing among a huge space of antecedently available innate primitives. The concept KANGAROO was not laying in wait in a system of representations available for selection by a Bayesian hypothesis testing mechanism, nor is it constructible by logical combination from antecedently available primitives.

Rey (2014) discusses the Margolis kind learning module, claiming

that it falls prey to Goodman's grue problem, just as Quinian bootstrapping does (see below). There are two answers to Rey's questions regarding constraints on induction in the Margolis kind learning module. First, as detailed above, there need be no induction. But, Rey asks, why are not kinds such as objects, animals, agents, Eastern grey kangaroos, kangadiles (kangaroos until year 2040, thereafter crocodiles), undetached kangaroo parts, or an infinitude of other kinds, possible glosses of SAME KIND AS THAT OBJECT, rather than the kind kangaroo? Why does the learner not form a concept of a particular individual (Oscar) instead of a kind?

Answering this question simply *is* an important part of the project understanding conceptual development. In the case of dedicated concept learning devices such as the object-kind learning device, the empirical project is specifying the constraints under which the system operates. That there is a dedicated kind concept acquisition device is an empirical discovery, and, like all learning mechanisms this one embodies strong constraints. It is a discovery that there is basic level in kind concepts, and it is a discovery that basic level kinds are privileged in kind concept learning (e.g., Rosch et al. 1976). It is a discovery that kind representations embody constraints derived from causal/functional analyses (see the work on psychological essentialism and the psychology of a causal/explanatory core to kind concepts: e.g., Gelman 2003, Keil 1989, Ahn and Kim 2000, Strevens 2000). And the existence and structure of systems of core cognition (in which the concepts AGENT and OBJECT are embedded), as well as innately supported systems of causal and functional analysis, are empirical discoveries, as is the fact that these constrain kind representations from early infancy (Carey 2009). These constraints do not rule out *ever* entertaining concepts for attended individuals. After all, some concepts that are not basic level are themselves innately manifest (e.g., AGENT) and are drawn upon as important parts of the constraints on the kind module. That is, AGENT is the content of a superordinate kind that constrains a newly formed basic level kind concept that falls under it. Others, such as subordinate and superordinate kinds, as well as stage and phase sortals like PUPPY or PASSENGER, are routinely manifested after basic level kind representations are formed (e.g., Hall and Waxman 1993). Still others are obviously entertainable (after all, Goodman and Quine did so, and we

all can join in). But these concepts simply are not the output of the dedicated basic level kind learning device discussed above. Furthermore, the child *can* also form a concept of a particular individual, even a newly encountered kangaroo. There is a dedicated learning mechanism for concepts of individuals, as well as for basic level kinds (but that is another story, one that has also been told; e.g., Belanger and Hall 2006). *Once* cognitive science has discovered the constraints under which actual learning devices operate, one can explore their possible outputs. The constraints posited are empirical proposals, falsifiable by demonstrations that they are easily violated. The empirical work strongly supports the existence of the basic level object kind learning module.

The basic level kind learning module creates new primitive concepts. Before a person has formed the concepts KANGAROO or SHOVEL, or concepts of any of infinitely many new kinds, he or she cannot think thoughts about the entities that fall under those concepts. This learning mechanism thus results in an increase in expressive power. However, like the cases of the dedicated learning mechanisms discussed in the ethology literature (those that yield representations of conspecifics, caretakers, the north star), there is an innately specified conceptual role for kind concepts, in this case given by the abstract concept KIND OF OBJECT and by the schemas of superordinate kinds embedded in core cognition and constructed theories that the learner assigns the new concepts to. Such already existing schema and conceptual roles are always part of the relatively easy route to new primitives.

8 The dual factor theory of representations with innate conceptual role

Dual factor theory applies straightforwardly to concepts in core cognition (AGENT, OBJECT...), indeed any concept with innate conceptual role and innate perceptual input analyzers that support identification of entities that fall under it. The innate perceptual input analyzers explain how symbols are causally connected to the entities they represent, and the innate conceptual role specifies the narrow content of the concept. In core cognition, and cases like the indigo bunting

representations of the azimuth, the innate conceptual role is never overturned—the narrow content of the representation of the north star that makes it a representation of north simply *is* the suite of sensori-motor computations supporting navigation it enters into.

The story for the Block/Macnamara/Margolis kind module is a little less straightforward. In concepts created by the kind learning device there are innate input analyzers that trigger the establishing of a kind representation (that identify objects with non-accidental structure) and that support the identification of superordinate schema provided by core cognition (KIND OF OBJECT, KIND OF AGENT...). These innate input analyzers are part of what provides the wide content of such concepts, as they trigger forming a representation of an entity in the world that is part of the wide content of the newly formed concept, as well as providing part of the causal connection between this wide content and the newly formed mental symbol. But there is no innate, un-overturnable, prepared conceptual role at the level of specific kinds. Even the initial superordinate schema the kind is subsumed under is revisable. However there is innate conceptual role for object kinds in general (i.e., given by psychological essentialism), and this specifies what sort of concept is in play and constrains its formal properties. This abstract conceptual role specifies part of the narrow content for kind concepts. As Block (1986) says, it determines the nature of the connection between symbols and the world, after a symbol is taken to be a symbol for an object kind.

9 The relatively hard route to new conceptual primitives

Quinian Bootstrapping is a learning mechanism that also creates new primitives, thus increasing the expressive power of the conceptual system. It differs from those learning mechanisms described above in that it did not arise through natural selection to acquire representations of a particular sort. Rather, it is one of the learning mechanisms that underlie the creation of representational resources that are discontinuous with (in the sense of being qualitatively different from, being locally incommensurable with, the representations of the same domain that were their input). It creates new conceptual roles, rather than merely creating new primitives for which there were prepared conceptual roles (as in the case in the easy route to

new primitives, see above). But once created, these new conceptual roles provide constraints on the concepts that will be learned, just as in the relatively easy route to new conceptual primitives.

TOOC takes a particular episode along the way to creating a representation of integers as a central worked example of conceptual discontinuity and of Quinian bootstrapping. I argue that this case involves an increase in expressive power, in that before the bootstrapping episode the child has no manifest concepts for natural numbers, and the process of construction of the first representations of new primitive concepts, those of a subset of the natural numbers, is not exhausted by defining them in terms of primitives antecedently available. Again, let me be clear. The increase in expressive power at stake here is an increase in the expressive power of manifest concepts available to the child. Obviously the total computational machinery available to the child has the capacity for this construction (what is actual is possible); just as the computational machinery of the child has the capacity to create representations of kangaroos in the easy route to new primitives.

Expressive power is a semantic/logical issue. Examples of questions about expressive power relative to number representations include whether arithmetic can be expressed in the machinery of sentential logic (provably no) and whether arithmetic can be expressed in the machinery of quantificational logic plus the principle that 1-1 correspondence guarantees cardinal equivalence (provably yes, if you accept Frege's proof). But the exploration of expressive power with such proofs is relevant to the question of how arithmetic arises in development *only* against empirically supported proposals for what the innate numerically relevant primitives are, and what form innate support for logic takes. If arithmetic can be derived from the resources of logic alone (with no numerical primitives), this is relevant to the question of the origin of arithmetic in ontogenesis *only* if the relevant logical resources are innate, and in a form that would support the relevant construction. If primitives with numerical content are needed as well (e.g., the principle that 1-1 correspondence guarantees cardinal equivalence, or the concepts ONE and SUCCESSOR), then one must account for how these arise in development. *TOOC* provides evidence that these numerical concepts are not part of the child's innate endowment, and that they arise only after the

bootstrapping episode in which the numeral list representation of number is constructed.

TOOC does not consider the form innate support for logic takes, and how logical resources arise in development. Indeed, I am acutely aware of this lacuna, and of its relevance to our understanding of numerical development. These questions have been the focus of research in my lab for the past four years, and will be so for the next decade at least. We do not yet have answers concerning the form innate support for logic takes. My current guess is that innate logic is largely implicit, embodied in computations, and that bootstrapping is needed before children create the logical resources needed for the mathematical construction of the integers from such primitives. After all, these constructions did not arise in mathematics until after 2000 years of development of formal logic. However, as I say below, my picture of the ontogenesis of concepts of integers would be falsified by the discovery of manifest representations with numerical content in addition to the three systems for which we already have empirical support.

Thus, I acknowledge that Fodor (2010), Leslie et al. (2007), Rey (2014), Rips et al. (2008), and others *could turn out to be* right (not that they provide a shred of evidence) that a full characterization of the manifest initial state will reveal expressive power sufficient to express arithmetic. If so, I would certainly back away from my claims about this bootstrapping episode increasing expressive power, saying that my studies concern how arithmetic capacities *actually* become manifest in ontogenesis. After all, the latter is actually my concern. I am quite certain that children do not construct arithmetic as Peano/Dedekind or Frege did, and I favor my bootstrapping story about what children actually do. But, if numerical or logical primitives are needed that themselves arise as a result of bootstrapping processes, then my claims of increases in expressive power stand.

At any rate, the actual process through which representations of integers arise is an existence proof of the possibility that bootstrapping *can* yield new primitives. The case study of the ontogenetic origin of integer representations illustrates all three major theses in *TOOC*: the existence of conceptually rich innate representations, conceptual discontinuity, and Quinian bootstrapping.

10 Core cognition of number (rich innate representational resources; *TOOC*, Chapter 4⁴)

Core cognition contains two systems of representation with numerical content: parallel individuation of small sets of entities in working memory models, and analog magnitude representations of number. Analog magnitude representations were briefly sketched in section 3 above. They are analog symbols of approximate cardinal values of sets. One signature of this system of number representation is that magnitudes are compared to one another on the basis of their ratios, and thus discriminability accords with Weber's law (discriminability is better the smaller the absolute value of the quantity) and exhibits scalar variability (the standard deviation of multiple estimates of a given quantity is a linear function of the absolute value of that quantity.) Analog magnitude representations of number have been demonstrated in many animals (rats, pigeons, non-human primates) as well as in humans from neonates to adults.

Analog magnitude representations are the output of paradigmatic perceptual input analyzers, but the analog magnitude symbols for number that are produced are conceptual in the sense of having rich central conceptual roles, including the many different arithmetical computations they enter into, and the fact that they are bound to (quantify over) many types of individuals (objects, events, auditory individuals).

A second system of core cognition with numerical content, parallel individuation, consists of working memory representations of small sets of individuals (three or fewer). The symbols in this system represent individuals (e.g., a set of 3 crackers is represented CRACKER CRACKER CRACKER, probably with iconic symbols for each cracker). Unlike the analog magnitude number representation system, parallel individuation/working memory is not a dedicated number representation system, nor are there any symbols that represent cardinal values (or any other quantifiers) in these models; there are only symbols for individuals. These models are used to compute total volume and area of the individuals, and are input into spatial and

⁴ The evidence for central claims in *TOOC*, along with citations of relevant literature, can be found in the chapters flagged throughout the current text.

causal representations. The numerical content in the system of parallel individuation is entirely implicit; the symbols in the models stand in 1-1 correspondence with individuals in the sets modeled. This is ensured by computations sensitive to spatiotemporal cues to numerical identity. The system must determine whether a given individual is the same one or a different one from a previously viewed individual to determine whether to add another symbol to the model. Further implicit numerical content is embodied in some of the conceptual roles these models enter into. More than one model can be entertained at any given time, and models can be compared on the basis of 1-1 correspondence to establish numerical order and equivalence. Importantly, this system of representation implicitly represents *one*. There is no explicit symbol with the content *one*, but the system updates a model of a set of one when a numerically distinct individual is added to it, yielding a model of a set of two (and ditto for sets of two and three), and the system similarly updates a model if individuals are removed from it. There is a strict upper limit to the number of individuals that can be held in working memory at any given time: 3 for infants. This set-size limit on performance contrasts with the ratio limit on performance that characterizes analog magnitude systems.

The parallel individuation system is perception-like in many ways, especially if the symbols for individuals are indeed iconic, as I suspect. Nonetheless the parallel individuation models themselves are conceptual in that they are held in a working memory system that requires attention and executive function, and enter into many further computations in support of rich central inferential processes (e.g., reasoning about the actions of agents upon objects, functional analyses, causal analyses, as well as quantitative computations).

Systems of core cognition are not the only innate resources relevant to conceptual development. *TOOC* assumes also early linguistic resources, but makes no attempt to specify their exact nature (a topic for another book). And, as commented above, the nature of logical resources available to infants and toddlers is virtually unstudied. Particularly relevant for number representations are linguistic representations that underlie the meanings of natural language quantifiers. Number marking in language (quantifiers, determiners, singular/plural morphology) requires representations of sets and individuals,

and provides explicit linguistic symbols with numerical content ‘a, all, some, most, many, few...’. *TOOC* reviews evidence that before age 2 children have mastered some of the basic syntax and semantics of natural language quantifiers, and that these linguistic structures provide important early constraints on the meanings of verbal numerals, via syntactic bootstrapping.

11 Conceptual discontinuity (*TOOC*, Chapter 8)

There are two steps to establishing discontinuities in development. The first, most important, step is characterizing the nature and content of symbols in successive systems of representation: Conceptual Systems 1 and 2 (CS1 and CS2). These characterizations allow us to take the second step: namely, to state precisely how CS2 is qualitatively different from CS1. With respect to numerical content, there are three CS1s: analog magnitude representations, parallel individuation, and natural language quantification.

The substantive claims in *TOOC* are that these three systems of representation exist, have been characterized correctly, and are the *only* representational systems with numerical content manifest in infancy and the toddler years. *TOOC*’s picture of number development would be falsified if evidence were to be forthcoming for innate numerical representations in addition to those described above, or different from them. Indeed, one aim of my current work on the logical resources of infants and toddlers is to search for such evidence.

CS2, the first explicit representational system that represents even a finite subset of the positive integers, is the verbal numeral list embedded in a count routine. Deployed in accordance with the counting principles articulated by Gelman and Gallistel (1978), the verbal numerals implicitly implement the successor function, at least with respect to the child’s finite count list. For any numeral that represents cardinal value n , the next numeral in the list represents $n + 1$.

CS2 is qualitatively different from each of the CS1s because none of the CS1s has the capacity to represent any integers. The new primitives are the concepts 1, 2, 3, 4, 5, 6, 7, the concepts that underlie the meanings of verbal numerals. Parallel individuation includes no summary symbols for number at all, and has an upper limit of 3 or

4 on the size of sets it represents. The set-based quantificational machinery of natural language includes summary symbols for quantity (e.g., 'some, all') and importantly contains a symbol with content that overlaps considerably with that of 'one' (namely, the singular determiner, 'a'), but the singular determiner is not embedded within a system of arithmetical computations. Also, natural language set-based quantification has an upper limit on the set sizes that are quantified with respect to exact cardinal values (i.e., TRIAL, along with, SINGULAR and DUAL). Analog magnitude representations include summary symbols for quantity that are embedded within a system of arithmetical computations, but they represent only approximate cardinal values, and their format is analog. There is no representation of exactly 1, and therefore no representation of $+1$. Analog magnitude representations cannot even resolve the distinction between 10 and 11 (or any two successive integers beyond its discrimination capacity), and so cannot express the successor function. Thus, none of the CS1s can represent 10, let alone 342,689,455.

As required by CS2's being qualitatively different from each of the CS1s that contain symbols with numerical content, it is indeed difficult to learn. American middle-class children learn to recite the count list and to carry out the count routine in response to the probe 'how many', shortly after their second birthday. They do not learn how counting represents number for another $1\frac{1}{2}$ or 2 years. Young two-year-olds first assign a cardinal meaning to 'one', treating other numerals as equivalent plural markers that contrast in meaning with 'one'. Some 7 to 9 months later they assign cardinal meaning to 'two', but still take all other numerals to mean essentially 'some', contrasting only with 'one' and 'two'. They then work out the cardinal meaning of 'three' and then of 'four'. This protracted period of development is called the 'subset'-knower stage, for children have worked out cardinal meanings for only a subset of the numerals in their count list.

Many different tasks, which make totally different information processing demands on the child, confirm that subset-knowers differ qualitatively from children who have worked out how counting represents number. Subset-knowers cannot create sets of sizes specified by their unknown numerals, cannot estimate the cardinal values of sets outside their known numeral range, do not know what set-size

is reached if 1 individual is added to a set labeled with a numeral outside their known numeral range, and so on. Children who succeed on one of these tasks succeed on all of them. Furthermore, a child diagnosed as a 'one'-knower on one task is also a 'one'-knower on all of the others, ditto for 'two'-knowers, 'three'-knowers and 'four'-knowers. The patterns of judgments across all of these tasks suggest that parallel individuation and the set-based quantification of natural language underlie the numerical meanings subset-knowers construct for numeral words.

Also consistent with the claim of discontinuity, studies of non-verbal number representations in populations of humans who live in cultures with no count list (e.g., the Piraha: Gordon 2004; Frank et al. 2008; the Mundurucu: Pica et al. 2004), and populations of humans in numerate cultures with no access to a count list (e.g., homesigners, Spaepen et al. 2011) show no evidence of any number representations other than the three CS1s.

In sum, the construction of the numeral list representation is a paradigm example of developmental discontinuity. How CS2 transcends CS1 is precisely characterized, and consistent with this analysis, CS2 is difficult to learn and not universal among humans.

12 Greater expressive power?

The above analysis makes precise the senses in which the verbal numeral list (CS2) is qualitatively different from those manifest representations with numerical content that precede it: it has a totally different format (verbal numerals embedded in a count routine), none of the CS1s with numerical content can express, even implicitly, an exact cardinal value over 4. But is the argument that the concepts for specific integers are new *primitives*, undefinable in terms of preexisting concepts using the combinatorial resources available to the child, actually correct? This argument, if correct, establishes the claim that acquiring the verbal count list representation of integers increases expressive power. As I comment in *TOOC*, this is on its face an odd conclusion. Integers are definable, after all, in terms of many different possible sets of primitives (e.g., 1 and the successor function, or the principle that 1-1 correspondence guarantees numerical equivalence plus the resources of quantificational logic).

At issue is whether logical combination underlies the transition from CS1 (core cognition of number) to CS2 (representations of verbal numerals that implicitly express the successor function). This is only possible if the capacity to represent integers is innate (e.g., if there is an innate representation of ONE and SUCCESSOR), or if integers are definable, by logical construction, from *manifest* innate primitives using *manifest* logical processes of conceptual combination. Whether acquiring integer representations increases expressive power simply is this question. Without a full characterization of the manifest combinatorial (logical) apparatus available to the child at the time the integers are constructed one cannot definitively answer the question of whether the child *could in principle* construct integer representations from innate resources, quite apart from the question of whether this is how the child *does* arrive at integer representations. But one can explore how the child actually *does* do so, and, in the remaining pages of this paper, I explain why I believe the process is *not* one of logical construction.

It's true that humans must ultimately be able to formulate concepts of integers using the explicit machinery of logic, enriched by whatever numerical concepts are necessary as well (what is actual is possible). But it is only after very long historical, and ontogenetic, developmental processes that the construction of integers in terms of logic or Peano's axioms is made. We simply do not know whether part of this process involved bootstrapping new logical representations as well as new numerical primitives.

13 A logical construction of the cardinal principle

Piantadosi et al. (2012) demonstrated that children could, in principle, construct a count list representation of the integers (at least up to 'ten') by conceptual combination alone, given the full general resources of logic (in the form of logical and set operations—if/then, set difference, plus lambda calculus, including the capacity for recursion), knowledge of the structure of the count list (its order), and four numerical primitives: the concepts SINGLETON, DOUBLET, TRIPLET, and QUADRUPLER (i.e., already manifest concepts of 1, 2, 3, and 4). Piantadosi et al. appeal to the literature on learning to count in support of the claim that these numerical concepts and a

representation of the count list are manifest at the time of the induction of the counting principles, but they merely assume—without evidence—that full general resources of lambda calculus and logic are available for the generation of hypotheses about what ‘one’, ‘two’, ‘three’, ‘four’, ‘five’...through ‘ten’ mean. They assume that children learn the meanings of the words ‘one’ through ‘ten’ from hearing words in cardinal contexts, through Bayesian enumerative induction. Thus, their model satisfies Fodor’s premises 1 and 2.

The model receives input in the form of sets with 1 to 10 items paired with the appropriate verbal numeral. It learns a function, in the language of lambda calculus, that allows it to answer the question ‘how many individuals?’ with the correct numeral. The model’s input reflects the relative frequency of verbal numerals in parental speech to children (i.e., ‘one’ is vastly more frequent than ‘two’, and so on.) Learning is constrained by limiting the combinatorial primitives that articulate hypotheses to be evaluated to those detailed above, by a preference for simpler hypotheses (i.e., shorter expressions in lambda calculus), and by a parameter that assigns a cost for recursion. After considering over 11,000 (!) different hypotheses composed from these primitives, the model learns to assign the words ‘one’ through ‘four’ to the concepts SINGLETON, DOUBLET, TRIPLET, and QUADRUPLER, and also (independently) learns a recursive cardinal principle knower function that assigns the numerals ‘one’ through ‘ten’ to sets of one through ten individuals. The recursive function tests whether the set in question (S) is a singleton, and if so, answers ‘one’. If not, it removes an element from S, and computes ‘next’ in the count list. It then applies the same singleton probe on the resultant set. If the answer is now yes, it outputs the numeral achieved by the ‘next’ function (i.e., ‘two’.) If not, it recursively repeats this step, stepping up through the count list and down through the set until a singleton results from the set difference operation.

The model matches, qualitatively, several details of children’s learning to count: children go through ‘one’-, ‘two’-, ‘three’- and ‘four’- knower stages, in that order, and depending upon the cost assigned to recursion, learn the CP-knower function after becoming ‘three’-knowers or ‘four’-knowers. Before the model learns the recursive CP-function, it has no way of knowing what numeral to apply to sets greater than 4, and in this sense Piantadosi et al. claim

a discontinuity in the model's knowledge of number word meanings. Thus, they claim for this model that it puts bootstrapping on a firm computational basis, as well as focusing on the logical resources actually needed for bootstrapping to succeed.

Piantadosi et al. assert that combination is the source of novelty. Therefore, in the current discourse, they are denying a genuine discontinuity. There is no change in expressive power—the manifest primitives (both numerical and logical) clearly can, in combination, express the cardinal meanings of 'one' through 'ten'. I will show why this model does not implement Quinian bootstrapping after I've discussed Quinian bootstrapping (see Rips, Asmuth and Bloomfield 2013, for an illuminating discussion). Here I simply want to acknowledge that, of course, depending upon the manifest concepts (both numerical and logical) actually available to the child, it certainly could be possible to learn the meanings of verbal numerals by constructing them from antecedently available concepts through logical combination.

The question that concerns me is how representations of integers *actually* arise in development. In what follows, I sketch a very different picture, one that does not rely on conceptual combination alone, and provide reasons to believe that this is the correct picture. My goal is to provide reasons to doubt that hypothesis formation by logical combination from primitives is the *only* source of new concepts.

14 Quinian bootstrapping

In Quinian bootstrapping episodes, mental symbols are established that correspond to newly coined or newly learned explicit symbols. The latter are initially placeholders, getting whatever meaning they have from their interrelations with other explicit symbols. As is true of all word learning, newly learned symbols must necessarily be initially interpreted in terms of concepts already available. But at the onset of a bootstrapping episode, these interpretations are only partial—the learner does not yet have any manifest concepts in terms of which he or she can formulate the concepts the symbols will come to express.

The bootstrapping process involves aligning the placeholder structure with the structure of existent systems of concepts that are

manifest in similar contexts. Both structures provide constraints, some only implicit and instantiated in the computations defined over the representations. These constraints are respected as much as possible in the course of the modeling activities, which include analogy construction. When the bootstrapping is under metaconceptual control, as is the case when it is being carried out by adult scientists, the analogical processes are explicit, and the fruitfulness of the analogies are monitored, and other modeling processes are also deployed, such as limiting case analyses, and thought experiments. Inductive inference is also often involved in bootstrapping, constrained by the actual conceptual structures in the process of being aligned.

In the case of the construction of the numeral list representation of the integers, the memorized count list is the placeholder structure. Its initial meaning is exhausted by the relations among the external symbols: they are stably ordered and applied to a set of individuals one at a time. 'One, two, three, four...' initially has no more meaning for the child than 'a, b, c, d...', if 'a, b, c, d...' were to be recited while attending to individuals one at a time.

The details of the subset-knower period suggest that the resources of parallel individuation, enriched by the machinery of linguistic set-based quantification, provide numerical meanings for the first few numerals, independently of their role in the memorized count routine. Le Corre and I (2007) proposed that the meaning of the word 'one' is represented by a mental model of a set of a single individual $\{i\}$, along with a procedure that determines that the word 'one' can be applied to any set that can be put in 1-1 correspondence with this model. Similarly 'two' is mapped onto a long term memory model of a set of two individuals $\{j, k\}$, along with a procedure that determines that the word 'two' can be applied to any set that can be put in 1-1 correspondence with this model. And so on for 'three' and 'four'. This proposal requires no mental machinery not shown to be in the repertoire of infants—parallel individuation plus the capacity to compare models on the basis of 1-1 correspondence. But those representations are enriched with the long-term memory models that have the conceptual role of assigning 'one', 'two', 'three', and 'four', to sets during the subset-knower stage of acquiring meanings for verbal numerals. We suggested that enriched parallel individuation might also underlie the set-based quantificational machinery

in early number marking, making possible the singular/plural distinction, and in languages that have them, dual and trial markers. The work of the subset-knower period of numeral learning, which extends in English-learners between ages 2:0 and 3:6 or so, is the creation of the long term memory models and computations for applying them that constitute the meanings of the first numerals the child assigns numerical meaning to.

Once these meanings are in place, and the child has independently memorized the placeholder count list and the counting routine, the bootstrapping proceeds as follows: The child must register the identity between the singular, dual, trial, and quadral markers and the first four words in the count list. In the course of counting the child notes (at least implicitly) the suspicious coincidence that the numeral reached when counting a set of 'three' is also the word a 'three'-knower takes to represent the cardinal value of that set. This triggers trying to align these two independent structures. The critical analogy is between order on the list and order in a series of sets related by an additional individual. This analogy supports the induction that any two successive numerals in the child's finite count list will refer to sets such that the numeral farther in the list picks out a set that is 1 greater than that earlier in the list.

In my earliest writings I characterized the induction made by 4-year-olds as yielding the first representations of integers. Let me be clear, as *TOOC* is, when the child has built the count list representation of the first 10 or so verbal numerals, the child does not yet have general representation of integers. There are many further bootstrapping episodes along the way to a representation of integers, two of which are discussed in *TOOC*—about 6 months after becoming CP-knowers, children construct a mapping between the count list and analog magnitude representations, yielding a richer representation of the meanings of verbal numerals (Chapter 9). Shortly thereafter, children abstract an explicit concept *NUMBER*, and explicitly induce that there is no highest number (Hartnett and Gelman 1998). And it is not until late in elementary school or even high school that children construct a mathematical understanding of division that allows them to reanalyze integers as subset of rational numbers (Chapter 9). All of these developments are along the way to richer and richer representations of integers. But without the construction of an integer

list representation of a finite subset of integers, which provides children with new primitive concepts for specific integers beyond four (e.g., 'seven' representing exactly seven) as well as providing new representations of 'one' through 'four' (in terms of their place in a count list, rather than only in terms of enriched parallel individuation) these further bootstrapping episodes never get off the ground.

This proposal illustrates all of the components of bootstrapping processes: placeholder structures whose meaning is provided by relations among external symbols, partial interpretations in terms of available conceptual structures, modeling processes (in this case analogy), and an inductive leap.

The greater representational power of the numeral list than that of any of the systems of core cognition from which it is built derives in part from creating a new representational structure—a count list—a new conceptual role—counting, and just *using it*. Much of the developmental process involves no hypothesis testing. Just as when the child learns a new telephone number (memorizes an ordered list of digits) and learns to use it in a procedure (dial, press buttons) that results in a ring and connection to Daddy, here the child learns an ordered list and procedure for applying it to individuals as one touches them one at a time. This new structure comes to have numerical meaning through the alignment of aspects of its structure with aspects of the structure of manifest number representations. These, in turn, have been built from set-based quantification (which gives the child singular, dual, trial, and quadral markers, as well as other quantifiers), and the numerical content of parallel individuation (which is largely embodied in the computations carried out over sets represented in working memory models with one symbol for each individual in the set). The alignment of the count list with these manifest meanings is mediated, in part, by the common labels (the verbal numerals) in both structures. At the end of the bootstrapping episode, the child has created symbols that express information that previously existed only as constraints on computations. Numerical content does not come from nowhere, but the process does not consist of *defining* 'seven' by conceptual combination of primitives available to infants. 'Seven' is genuinely a new primitive, the meaning of which is provided in part by its conceptual role in a new conceptual structure.

With this characterization in hand, one can see why the Piantadosi et al. (2012) model does not implement a Quinian bootstrapping process. There are three theoretically important differences between Quinian bootstrapping and a model that formulates hypotheses at random by explicit conceptual combination from 15 primitives, one numeral at a time, and then uses Bayesian induction to evaluate them. First, although, like Piantadosi et al., I assume that children have representations with the content SINGLETON, DOUBLET, TRIPLET, QUADRUPLER, before the child induces the cardinal principles, the numerical content of these representations is carried by enriched parallel individuation, and is merely implicit until the child constructs the relevant structures. On this proposal there are no manifest summary discrete symbols for these concepts. The first explicit symbols are 'one', 'two', 'three' and 'four' and their meanings are not already existing primitives SINGLETON, DOUBLET, TRIPLET, QUADRUPLER. Similarly, the representations that underlie the meaning of 'seven', after the cardinal principle induction, are largely implicit in the procedures of the count routine, not explicitly defined in a language of thought. Second, the meanings of numerals in the Piantadosi model are learned entirely independently from each other. That is, children could, in principle, compose the recursive definition of numerals first, without ever going through 'one', 'two', 'three', and 'four'-knower stages. In Piantadosi's model, although the primitive SINGLETON plays a role in the cardinal principle function, knowing the meaning of 'one' (expressing the innate primitive SINGLETON) plays no role in learning the meanings of other numerals nor learning the cardinal principle underlying how counting expresses number. In Quinian bootstrapping, the structure created by interrelations of the newly learned words, plus their partial meanings from initial mappings to prelinguistic thought, play an essential, constitutive role in the learning process. Thirdly, and relatedly, the Quinian bootstrapping story takes seriously the question on the source of constraints on the learning process. It empirically motivates its claims of the exhaustive set of primitives with numerical content, (the three CS1s), and provides evidence for syntactic bootstrapping as an account for how the child breaks into the meanings of the first numerals. As Rips et al. (2013) point out in their illuminating discussion of the Piantadosi model, this model does not

provide an account for how the hypothesis space is conveniently limited to just the 15 numerically relevant primitives it randomly generates hypotheses from. The child has much other numerically relevant knowledge at the time of the CP induction. If that knowledge were included in the set of primitives, the hypothesis space created by random combination from the primitives would explode beyond the already entirely unrealistic 11,000 hypotheses considered and rejected by the model. If numerically irrelevant primitives are included (how does the child decide which primitives are relevant?), the problem would quickly become entirely intractable.

In sum, Quinian bootstrapping is very different from the Piantadosi logical combination model, but which model provides better insight into how children actually learn how counting represents number? Two recent animal studies clarify the nature of bootstrapping, allowing us to see that it does not involve hypothesis testing over a huge space of existing concepts, nor does it involve logical combination of primitives. These studies also increase the plausibility that young children have the computational resources to engage in Quinian bootstrapping.

15 Animal models

In *TOOC* I speculated that Quinian bootstrapping might well be a uniquely human (depending upon external explicit symbols as it does), and thus might provide part of the explanation for the uniquely human conceptual repertoire. Since then, two studies have convinced me that other animals have the capacity for Quinian bootstrapping.

15.1 Alex

The first study (Pepperberg and Carey 2012) drew on explicit numerical representations created by Alex, an African grey parrot, who had been trained by Irene Pepperberg for over 30 years. He had a vocabulary of over 200 words, including object labels, color words, relational terms such as ‘same’, and several other types of labels. Alex had been taught to produce the words ‘three’ and ‘four’ in response to ‘how many x’s’ for sets of 3 and 4 respectively. During

this initial training, Alex was also shown mixed sets of objects (e.g., 4 blue balls, 5 red balls, and 3 yellow balls), and asked, for example, 'what color three,' responding 'yellow.' In other words, he was first taught to produce and comprehend 'three' and 'four' as symbols for cardinal values 3 and 4. After this training was in place, he was similarly taught to produce and comprehend 'two' and 'five' as symbols for cardinal values 2 and 5. And then 'one' and 'six' were added to his repertoire.

We do not know what non-linguistic numerical representations underlay these explicit numeral representations, because we do not know the sensitivity of Alex's analog magnitude representations or the set size limit of his parallel individual system. Analog magnitude representations themselves could have done so, or both parallel individuation and analog magnitudes could have been drawn upon. As he is a non-linguistic creature, he doesn't have the resources of set-based quantification that is part of the language acquisition device to draw upon. What the quantificational resources of non-linguistic thought are has not been studied, but Alex clearly had the capacity to selectively attend to small sets and evaluate whether any given set had a cardinal value of 'one' through 'six'.

After he had a firm understanding of the cardinal meanings of 'one' through 'six', Pepperberg taught him to label plastic tokens of Arabic numerals '1, 2, 3, 4, 5' and '6', with the words 'one' through 'six' respectively. Arabic numerals were never paired with sets of individuals. The only connection between Arabic numerals and set sizes was through the common verbal numeral (e.g., 'two' for '2' and 'two' for a set of 2 individuals.) He quickly learned to produce and comprehend the verbal numeral labels for the Arabic numerals. Then with no further training, Pepperberg posed him the following question for each pair of Arabic numerals between '1' and '6': 'Which color bigger?' He was to choose, for example, between a blue '3' and a red '5', the plastic Arabic numeral tokens being the same size and the correct answer being 'red'. He succeeded at this task when first presented it; it required no further training. Not only had he not been given any positive evidence that '2' refers to a cardinal value, the only context in which he had answered questions about 'bigger' and 'smaller' previously was in with regards to physical size (i.e., 'which color bigger' of two objects that differed in size).

Please dwell on this finding. It must be that the common labels (e.g., 'two') had allowed him to connect a representation of the Arabic digits (e.g., '2') with the cardinal values (e.g., 2), and it must be that the intrinsic order in his nonverbal representations of cardinal values allowed him to say which Arabic numeral was bigger or smaller than which others. Although Alex had never been taught a count list (and had been taught the cardinal meanings of numerals in the order 'three/four', 'two/five' and finally 'one/six'), by the time we began our study Alex could produce and comprehend the words 'one' through 'six' as labeling both cardinal values of sets and Arabic digits, and could use the intrinsic order among set sizes to order his verbal numerals.

We were thus in a position to teach Alex to label Arabic numerals '7' and '8', 'seven' (pronounced by him 'sih-none' and 'eight' respectively). This training took about a year, and during it he had no evidence that '7' or '8' were numerals. He was then taught that '6' is a smaller number than '7', which in turn is a smaller number than '8', by posing the 'which color number bigger/smaller' task, giving him feedback if he guessed wrong. This was the first evidence he had that '7' and '8' are numerals, as are '1' through '6'. It took only a few hours to train him to answer which color number bigger or which color number smaller for each of the pairs: '6/7', '6/8' and '7/8'. After he had reached criterion on this task he was probed which color number bigger and smaller for each pair of numerals '1, 2, 3, 4, 5, 6' with respect to '7' and '8', and succeeded at this task with no further training. Thus, at this point he knew that '7' and '8' are verbal numerals, labeled 'sih-none' and 'eight' respectively, and he knew that '8' is a bigger number than '1' through '7' and '7' is a bigger number than '1' through '6'. Importantly, he had never been given any information about which cardinal values 'sih-none/7' and 'eight/8' referred to.

The question of this study was whether he would make the (wildly unwarranted) induction that 'sih-none/7' expresses cardinal value 7 and 'eight/8' expresses cardinal value 8. He did. The very first time he was asked to label a set of seven objects 'how many treats?' he answered 'sih-none' and the first time he was asked to label a set of eight objects 'how many treats?' he said 'sih-none' and immediately self corrected to 'eight'. Over a two week period he was asked

to label sets of different sizes. These questions were probed by many different experimenters, only a few questions a day, intermixed with many other questions currently under study, concerning visual illusions and many other things. He performed better than chance producing both 'sih-none' and 'eight' ($p < .01$ in each case). He was also given comprehension trials, (e.g., 'what color seven' and 'what color eight', probed with 3 sets or either 6, 7, 8, 9, or 10 colored blocks), and got 11 of 12 correct ($p < .01$). Thus, Alex had inferred the cardinal meanings of 'eight' and 'seven/sih-none' from knowledge of the cardinal meanings of 'one' through 'six' and from the fact that six is a smaller number than seven and seven is a smaller number than eight.

The Piantadosi model could not possibly apply here. This learning episode did not involve hypothesis confirmation. Alex never got any feedback as to whether his answers were correct. Nor was he ever given the pairings between 'seven (sih-none)' and sets of 7 and 'eight' and sets of eight that constitute the data for the Piantadosi model. Alex *must* have made an inductive inference based on the meanings of numerals he already had constructed. Given that his knowledge of the use of numerals was exhausted by just a few procedures they entered into (answering questions about set size and numerical order, labeling cardinal values of sets and labeling Arabic numerals), and by the mappings he had already made between representations of sets, verbal and Arabic numerals, his induction was subject to strong constraints. He clearly had not searched through a vast unconstrained hypothesis space specified by logical combination of all 250 or so concepts he had that were lexicalized (or even a larger set of conceptual primitives he may manifest). As mentioned, this induction was wildly unwarranted; what he had been taught was consistent with '7' referring to any set size greater than '6' and with '8' referring to any set size greater than whatever '7' refers to. But in his 30 years of working with numerals, they had been introduced as related by +1 ('three' vs. 'four', then 'two' and 'five', and then 'one' and 'six' added to his repertoire in turn). His induction was not mathematically or logically warranted, but it was sensible, given his actual experience with numerals. So too is the child's.

Piantadosi et al. might reply that Alex may have made the leap to CP knower, having engaged in the same conceptual combination process as hypothesized by their model that children do, during the

period of learning where he was taught ‘one’ through ‘six’. In that case, the induction he made here was that ‘seven’ and ‘eight’ were the next two numerals, in that order, in the relevant list after ‘six’. This is also not possible, because Alex lacked an essential set of primitive functions for the Piantadosi model during this earlier period: namely, he did not have a count list. He was never taught a list, *per se*, nor never taught to count. Thus he could not form any generalizations carried by the function *Next* applied to a count list. He wasn’t even taught the numerals in numerical order (remember he learned first ‘three’ and ‘four’, then ‘two’ and ‘five’ and finally ‘one’ and ‘six’). It’s true he explicitly knew his numerals were ordered, but that order had to be derived from the intrinsic order of cardinal values that were expressed by numerals and could not have been part of the source of the mapping between numerals and cardinal values. That order was not carried by a count routine and a memorized ordered list. Further insight into the process of learning Alex was more likely engaged in is provided by a recent study of rhesus macaques.

15.2 Rhesus macaques

Livingstone et al. (2009) taught four juvenile male rhesus macaques (1 year old at beginning of training), to choose the larger of two dot arrays, or to choose a symbol that came later in an arbitrary list. The dot arrays varied between 1 and 21 dots, and the arbitrary list of symbols was ‘1, 2, 3, 4, 5, 6, 7, 8, 9, X, Y, W, C, H, U, T, F, K, L, N, R’. The monkeys were trained on the dot arrays and on the symbol list on alternate days. Training in both cases involved giving the monkey a choice between two stimuli (e.g., 2 dots and 7 dots, or ‘2’ and ‘7’) on a touch screen. When the monkey touched one of the arrays, he was rewarded with the number of pulses of juice or water that corresponded to his choice. Thus he was always rewarded, but got bigger rewards for picking the larger dot array or the symbol later in the list. The monkeys learned to pick the stimulus that led to the larger reward with both stimuli sets, and were extremely accurate with both types of stimuli, making errors only for closely adjacent values.

There were two extremely interesting results that emerged from this study. First, with no training, the first time monkeys were given

a choice between dot arrays and symbols (e.g., 4 dots and '7'), they chose the stimulus that would lead to the larger reward. That is, they had automatically integrated the two predictors of quantity of liquid—dot arrays and discrete symbols ordered in a list). Clearly this integration had to be mediated by the fact that the dot array and discrete list tasks established a common context (same testing chamber, same dependent measure of touching one of two stimuli on a screen), and the outcomes predicted were from the same scale of quantities of liquid. Still, they had integrated them. This is the structural alignment process drawn upon in bootstrapping.

Second, when making a choice between dot arrays, the noise in choices among large sets (e.g., 19 vs. 21) was greater than that between smaller sets (e.g., 9 vs. 11 or 3 vs. 5). In fact, the choices showed scalar variability, the marker of analog magnitude values (see above). But errors in choosing values on the ordered list of discrete symbols did not increase as the list got longer. Livingstone et al. interpreted this difference as showing that the mapping from dot arrays to liquid quantity showed scalar variability, whereas the mapping from the list to hedonic value was linear. A more likely interpretation is that the mapping, during learning, reflected recognizing the relevance of each type of order (order among set sizes in analog magnitude representations of number of dots, and linear order in an arbitrary list) and inducing a rule that one should pick the stimulus later in each ordering. It's analog magnitude representations of dots that showed scalar variability, and the representations of the linear order in the list that did not. It's true that some mapping between each ordering and quantity of liquid was constructed in the process, because the two orderings were integrated. But if choosing between predicted quantities of liquid underlay each choice, both tasks should have shown scalar variability, since quantity of liquid is represented with an analog magnitude value. I suggest that the structure of an ordered list of symbols is a linear order, supported by the discriminability of each symbol from each other, and this order directly determined choice once the task was learned. This structure, after being constructed, was alignable with the intrinsic order of representations of quantity of liquid, and then with the other predictor of quantity of liquid (dot arrays). This is structurally the same as the alignment between an ordered list and analog magnitude

representations of number achieved some 6 months after children have become cardinal principle knowers.

Livingstone's rhesus macaques did not induce the cardinal meaning of a new symbol from its place in a count list, but nonetheless they exhibited several components of the extended bootstrapping process that supports children's (and Alex's) doing so. They did build a representation of an ordered list (21 elements long!) and align it with a representation that was itself intrinsically ordered. Also, they automatically aligned two different ordered representations (the list, the dot arrays) which were separately aligned to quantity of liquid. Clearly, finding the structural correspondence between an ordered list and increasing magnitude (whether that magnitude is a number or a continuous variable like quantity of liquid) is a natural computation, at least for primates.

15.3 Conclusions concerning the nature of bootstrapping

As the historical examples discussed in *TOOC* make clear, bootstrapping episodes are often under metaconceptual control; the scientist is consciously engaged in exploring mappings between mathematical structures and physical/biological/psychological phenomena. But as the above examples make clear, metaconceptually explicit hypothesis testing and modeling procedures are not necessary.

I now turn to the questions of whether the representations achievable by bootstrapping should be thought of as part of a preexisting hypothesis space, or otherwise as a process of formulating and confirming hypotheses in terms of concepts that are logical constructions from primitives in a preexisting hypothesis space.

Prior to the bootstrapping processes, neither children, nor Alex, nor rhesus macaques have any representations for exact cardinal values outside of the range of parallel individuation. A representation of 341,468, or of 10, does not exist in some preexisting hypothesis space ready to become manifest. Nor is a representation of 7 constructed by conceptual combination of innate primitives. Of course the child and Alex and the rhesus macaques, must have the capacity to represent linear order, and to construct a mapping between different ordered representations, but this process does not involve constructing definitions. Some of the learning processes involved in

the extended episode of bootstrapping are certainly not hypothesis testing (e.g., memorizing the ordered list of numerals), and some are subpersonal (as Shea (2011) put it ‘not explainable by content’; see also Strevens’ (2009) proposal that introjection involves subpersonal processes). That is, the connection of the ‘three’ in the count list with the ‘three’ of enriched parallel individuation is most probably mediated simply by the shared label and associative machinery, just as Alex’s aligning of his representations of verbal numerals, set sizes, and Arabic numerals is based first on common labels, which then supports ordering them according to the intrinsic order among cardinal values within AM and parallel individuation systems of representations. Similarly, the rhesus’ aligning of an ordered list of 21 discrete symbols with set sizes from 1 to 21 depends upon shared associations with quantities of liquid. Such alignment processes are not processes of logical combination (although logical combination is involved in building the placeholder structures). Also, Alex never got any feedback regarding the pairing of ‘seven’ and ‘eight’ with cardinal values, so no hypothesis confirmation or Bayesian enumerative induction was involved. I conclude that Quinian bootstrapping yields new primitives in this case, representations of integers embedded in a count list, and is a learning mechanism that does not conform to Premises 1 and 2 of Fodor’s argument.

16 Critiques of Quinian bootstrapping

Rey (2014), Fodor (2010), and Rips et al. (2013) deny Quinian Bootstrapping is a learning mechanism that can increase expressive power by creating new primitives not laying in wait. They deny that Quinian bootstrapping actually creates new primitives. It may create new concepts, but they are not primitives; they must be constructible by logical combination from others. Specific versions of the challenges include (1) analogy cannot create new representational resources, as analogies require alignable structures antecedently, (2) the induction the child makes requires an antecedent appreciation of the successor function, and (3) the bootstrapping proposal fails to confront Goodman’s grue problem, the problem of constraints on induction. As I hope is already clear, I believe all of these challenges to be off the mark.

With respect to the challenge that analogy requires already available representations to be aligned, I agree. The bootstrapping process is an extended one. The new representational resource is not created at the moment of the analogy and the induction alone. By the time of the induction of the counting principles, the child has indeed created the alignable structures needed for the limited induction he/she makes, just as Alex had. In the case of the child these structures are, by hypothesis, the count list and representations of the cardinal values of the numerals 'one' through 'four' supported by enriched parallel individuation. The whole process begins with the innate numerical resources (the CS1s described above), the enrichment of parallel individuation during the subset-knower stage, and the creation of the meaningless placeholder structure. Of course one needs both structures to align them. My account of the bootstrapping process specifies the origin of each structure and shows what new arises from their alignment.

I also don't agree with the second critique, that to notice sets of two differ from sets of three by a single individual, one must already represent the successor function. All one must be able to do is subtract 2 individuals from 3 individuals, and 1 individuals from 2 individuals, computations that both parallel individuation and analog magnitude representations support. The successor function, in contrast, generates an infinite series of cardinal values, whereas the knowledge the child has is initially restricted to the relations among sets of one, two, three and four (because of the set size limit on parallel individuation and the sensitivity of analog magnitude representations being limited to 3:4 or 4:5 among young preschoolers). But of course, without the capacity to subtract 2 individuals from a set of 3 individuals, and 1 individual from a set of two individuals, yielding a single individual in each case, the child could not make the induction concerning how his or her short count list works. I do not deny this knowledge must be in place for the induction; rather I present *evidence* that it is, including *how* it is (within the system of enriched parallel individuation in the case of children's learning to count), and *evidence* that precisely that induction separates subset-knowers from cardinal principle-knowers. Again, one must consider the format and computational roles of the actual representations involved. The induction the child most probably makes is that when you add

an individual to a set for which you would reach numeral N when applying the count routine, if you count the resulting set, you will reach the next word on the count list. This is not yet the successor function, and certainly doesn't presuppose the successor function.

Turning to the heart of Rey's and Rips et al.'s criticism: that I failed to take on Goodman's new riddle of induction. Rips et al.'s extended example of a possible induction consistent with the data children have available at the time of inducing the counting principles is modular arithmetic. They ask: why do children not hypothesize that the counting sequence begins at 1 again after reaching some value (e.g., 10, in a mod 10 system). That is, why do they not consider the hypothesis that the list cycles, just as 'Monday, Tuesday, Wednesday, ... Sunday, Monday...' does. Rey asks why children do not take 'two' to be a proper name for a set, or any of a myriad other hypotheses. There are, of course, an infinite number of hypotheses consistent with any finite set of data. Human inductive inference is profligate; so too, apparently, is parrot inductive inference. Accounting for the constraints on induction is everybody's problem. This paper has been an extended response to that critique. One place both writers go wrong is closely related to the view of possessed concepts as a vast hypothesis space, laying in wait to become manifest. If this were right (think Piantadosi et al.), the issue of constraints on induction would indeed be trenchant. As I have repeatedly said, any actual learning mechanism imposes constraints on what can be learned. Thus, part of the project of exploring an actual learning mechanism is studying what constraints are imposed by it, including constraints on induction. Of course children *could* learn a modular arithmetic (as adults can), but once integrated with analog magnitude representations, their actual hypotheses about meanings of numerals are constrained by the structure of the analog magnitude system (which extends open-endedly toward higher values), and constraints that the same words do not apply to discontinuous regions of it. Induction, in this case, is constrained by the only three systems of representations with numerical content (parallel individuation, analog magnitude representations, and natural language quantification) manifest at the time of learning.

One understands the constraints on the inductions made by 3-year-olds and by Alex by attending to the extremely limited

contexts in which these inductions (and most inductions) are actually drawn (think Alex and the rhesus macaques, as opposed to the model of Piantadosi et al., selecting among over 11,000 hypotheses consistent with the data it has received, where that large hypothesis space has been artificially constrained). The induction made during the hypothesized bootstrapping episode is constrained by the structures being aligned, and their very local conceptual roles. The scientific work involved in understanding episodes of Quinian bootstrapping is characterizing those structures, showing how they arise, and showing what new is achieved by aligning them.

17 A dual factor theory of bootstrapped concepts

Section 8 argued that dual factor theory straightforwardly applies to concepts that are easily acquired, for they are supported by innate conceptual roles that are never overturned (partially determining narrow content), and by innate perceptual input analyzers that guarantee a causal connection between entities in the world and the symbols that refer to them (partially determining wide content).

Chapter 13 of *TOOC* argues that dual factor theory is also needed to understand the nature of concepts that are the output of the bootstrapping episodes that underlie the origin of concepts that are hard to attain. Space does not permit a full discussion of this issue here. Briefly, newly coined concepts are ultimately mapped to antecedent ones that were supported by innate conceptual roles and innate input analyzers, and they inherit their wide content from that of those antecedent concepts. The placeholder structures in terms of which new concepts are introduced consist of interrelations among new concepts directly represented in an external language, not yet mapped to any already existing concepts that play any role in thought or refer to anything in the world. That is, they have *only* conceptual roles to provide their content. Bootstrapping proceeds by mapping these newly coined symbols to related symbols that are already interpreted. This process is often mediated by shared labels, but requires changes within the antecedently represented concepts, changes effected by aligning the two structures through modeling processes such as analogical mapping.

In *TOOC* (Chapter 13) I considered whether any of the conceptual

roles that play such an important role in this process determine the content of the final representations, given that they are all up for revision (and indeed, are revised in every episode of bootstrapping). The issue is that conceptual role has many roles to play in a full theory of concepts that do *not* specify narrow content, such as underlying inferences and being part of the sustaining mechanisms that connect concepts to their referents. The challenge to a dual factor theory is specifying which aspects of conceptual role, if any, actually determine content.

The proposal I made in *TOOC* was that the conceptual role that exhausts the meaning of the terms introduced in newly coined placeholder structures, and that constrains the structural alignment process through which these terms come to have wide content, is part of narrow content. But how can this be so, given that the relations expressed in placeholder structures are not analytic, but rather fall under the assumptions of psychological essentialism, and thus are assumed to be (and are) up for revision? The solution, I suggested, is to take seriously the relation between ancestor and descendant concepts in cases of true conceptual change (as opposed to cases of belief revision). Narrow content is that part of conceptual role that was part of the initial placeholder structure, or the aspects of conceptual role that led to changes at the level of individual concepts (differentiations, coalescences, changes in conceptual core) in the descendants of that initial placeholder structure.

18 Conclusions

As has long been recognized, a theory of concepts must include an account, at least in principle, of how it is possible that they are acquired, both over historical time and in ontogenesis. This problem has largely been ignored in the psychological literature on concepts within cognitive psychology. I have argued here that taking this problem seriously constrains our understanding of what concepts are. There are two broad routes to concept acquisition: the easy route that underlies episodes of fast mapping and the hard route that underlies conceptual discontinuities, and requires bootstrapping. The lesson that emerges from considering the two cases side by side is the crucial contribution of conceptual role in determining content. In

the easy cases, there are innate conceptual roles for the new concepts to play (NORTH in the night sky has an innate role to play in Bunting navigation; kind concepts are supported by an innate schema within the constraints of psychological essentialism). The hard cases differ from these in that there is no innate conceptual role for the new primitives, the new inferential role and the primitives that fill those roles must be co-constructed. The bootstrapping process includes constructing new placeholder structures whose symbols get meanings entirely in terms of their interrelations with each other, and this conceptual role then comes to have wide content through modeling processes that connect it to antecedently available representations. It is not a hard sell for psychologists to consider that inferential role must have a role to play in individuating concepts and specifying their content. Considerations of acquisition show both how deeply this is so, and provide suggestive evidence concerning the questions of which aspects of conceptual role might be content determining.

19 New directions

There is much work to be done, both on what I am calling the easy cases of concept acquisition and on what I am calling the hard cases. But here I want to draw attention to an urgent problem in this discourse that is virtually unstudied—specifying what form innate support for logic takes. We cannot evaluate Premise 2 of Fodor's argument without knowing this; we cannot know whether later developing concepts can be built from earlier available primitives by straightforward conceptual combination without this. One of the deepest issues in cognitive science is at stake. Many hold (e.g., Bermudez 2007; Penn et al. 2008) that non-human animals do not have a logic-like language of thought formulated over language-like representations of propositions, and many have suggested that these arise in development only upon learning natural language. Others (e.g., Braine and O'Brien 1998; Crain and Khlentzos 2010; Fodor 1975) hold that it is obvious that non-human animals have such representational capacities, and that babies could not learn language without it. Actually, it is not obvious one way or the other. It is possible that the capacity for logic-like conceptual combination may be part of the evolved capacity for human language and that it emerges in ontogen-

esis only in the course of language acquisition. More radically, it is possible that logical content is initially embodied only in computations defined on explicit representations, like the numerical content of parallel individuation, and that bootstrapping is needed to yield meanings of language-like symbols for logical connectives.

TOOC speculated that the format of representation of all core cognition systems is iconic, and provided evidence for this in the case of core cognition of number (both AM and PI representations). But systems of core cognition do not exhaust the innate representational repertoire. At the very least there are perceptual representations as well, and *perhaps* abstract representations of relations (e.g., CAUSE, SAME). It is less plausible that the format of these latter types of representations is iconic. Furthermore, it is completely unstudied whether infants have mental representations in their language of thought with the content of logical connectives, such as AND, OR, or NOT, but if there are, it is certain that their format of representation is not iconic. There is simply no research on logical symbols and reasoning schema in infancy using the productive methods of modern studies of infant cognition. There should be.

Susan Carey

Henry A. Morss Jr. and Elizabeth W. Morss Professor of Psychology
Psychology Department
Harvard University
33 Kirkland Street
Cambridge, Mass 02138
scarey@wjh.harvard.edu

References

- Ahn, W., and Kim, N. S. 2000. The role of causal features in categorization: An overview. In *Psychology of Learning and Motivation*, 40, ed. by D. L. Medin. New York: Academic Press, 23-65.
- Bélanger, J., and Hall, D. G. 2006. Learning proper names and count nouns: Evidence from 16- and 20-month-olds. *Journal of Cognition and Development* 7, 45-72.
- Bermúdez, J. L. 2007. Thinking without words: An overview for animal ethics. *The Journal of Ethics* 11(3): 319-335.
- Block, N. J. 1986. Advertisement for a semantics for psychology. In *Midwest studies in philosophy*, ed. by P. A. French. Minneapolis: University of

- Minnesota, 615-678.
- Braine, M. D. S., and O'Brien, D. P. (Eds.) 1998. *Mental Logic*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Carey, S. 1978. The child as word learner. In *Linguistic Theory and Psychological Reality*, ed. by J. Bresnan, G. Miller and M. Halle. Cambridge, MA: MIT Press, 264-293.
- Carey, S. 1982. Semantic development, state of the art. In *Language Acquisition, State of the Art*, ed. by L. Gleitman and E. Wanner. Cambridge: Cambridge University Press, 347-389.
- Carey, S. 1985. *Conceptual Change in Childhood*. Cambridge, MA: Bradford Books, MIT Press.
- Carey, S. 2009. *The Origin of Concepts*. New York: Oxford University Press.
- Crain, S. and Khlentzos, D. 2010. The logic instinct. *Mind and Language* 25(1): 30-65.
- Dehaene, S. 1997. *The Number Sense*. New York: Oxford University Press.
- Dretske, F. 1981. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Emlen, S. T. 1975. The stellar-orientation system of a migratory bird. *Scientific American* 233: 102-111.
- Fodor, J. A. 1975. *The Language of Thought*. Cambridge, MA: Harvard University Press.
- Fodor, J. A. 1980. On the impossibility of acquiring "more powerful" structures: Fixation of belief and concept acquisition. In *Language and Learning*, ed. by M. Piatelli-Palmerini. Cambridge, MA: Harvard University Press, 142-162.
- Fodor, J. A. 1998. *Concepts: Where cognitive science went wrong*. New York: Oxford University Press.
- Fodor, J. A. 2008. *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.
- Fodor, J. A. 2010, October 8. Woof, woof [Review of the book *The Origin of Concepts*, by S. Carey]. *Times Literary Supplement*, 7-8.
- Frank, M. C., Everett, D. L., Fedorenko, E., and Gibson, E. (in press). Number as a cognitive technology: Evidence from Piraha language and cognition. *Cognition* 108: 819-824.
- Gallistel, C. R. 1990. *The Organization of Learning*. Cambridge, MA: MIT Press.
- Gallistel, C. R., Brown, A., Carey, S., Gelman, R., and Keil, F. 1991. Lessons from animal learning for the study of cognitive development. In *The Epigenesis of Mind: Essays in Biology and Cognition*, ed. by S. Carey and R. Gelman. Hillsdale, NJ: Erlbaum, 3-36.
- Garcia, J., Kimeldorf, D. J., and Koelling, R. A. 1955. Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science* 122(3160): 157-8.
- Gelman, R. and Gallistel, C. R. 1978. *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.
- Gelman, S. A. 2003. *The Essential Child: Origins of essentialism in everyday thought*.

- New York: Oxford University Press.
- Gordon, P. 2004. Numerical cognition without words: Evidence from Amazonia. *Science* 306(5695): 496-499.
- Hall, D. G. and Waxman, S. R. 1993. Assumptions about word meaning: Individuation and basic-level kinds. *Child Development* 64: 1550-1570.
- Hartnett, P., and Gelman, R. 1998. Early understandings of numbers: Paths or barriers to the construction of new understandings? *Learning and Instruction* 8(4): 341-374.
- Keil, F. C. 1989. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.
- Keleman, D. and Carey, S. 2007. The essence of artifacts: Developing the design stance. In *Creations of the Mind: Theories of Artifacts and Their Representation*, ed. by E. Margolis and S. Lawrence. New York: Oxford University Press, 212-230.
- Kripke, S. 1972/1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Leslie, A. M., Gallistel, C. R., and Gelman, R. 2007. Where integers come from. In *The Innate Mind: Foundations and Future*, ed. by P. Carruthers, S. Laurence, and S. Stich. Oxford: Oxford University Press, 109-138.
- Levin, B. and Pinker, S. 1991. *Lexical and Conceptual Semantics*. Cambridge, MA: Blackwell.
- Livingstone, M. S., Srihasam, K., and Morocz, I. A. 2009. The benefit of symbols: monkeys show linear, human-like, accuracy when using symbols to represent scalar value. *Animal Cognition* 13: 711-719.
- Macnamara, J. 1986. *Border Dispute: The place of logic in psychology*. Cambridge, MA: MIT Press.
- Margolis, E. 1998. How to acquire a concept. *Mind and Language* 13: 347-369.
- Margolis, E. and Laurence, S. 2011. Learning matters: The role of learning in concept acquisition. *Mind and Language* 26: 507-539.
- Miller, G. A. 1977. *Spontaneous Apprentices: Children and Language*. Seabury Press.
- Miller, G. A. and Johnson-Laird, P. N. 1976. *Language and Perception*. Cambridge, UK: Cambridge University Press.
- Murphy, G. 2002. *The Big Book of Concepts*. Cambridge, MA: MIT Press.
- Penn, D. C., Holyoak, K. J., and Povinelli, D. J. 2008. Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences* 31(2), 109-129.
- Pepperberg, I. and Carey, S. 2012. Grey Parrot number acquisition: The inference of cardinal value from ordinal position on the numeral list. *Cognition* 125: 219-232.
- Piantadosi, S. T., Tenenbaum, J. B., and Goodman, N. D. 2012. Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition* 123: 199-217.
- Pica, P., Lemer, C., and Izard, V. 2004. Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306(5695): 499-503.
- Prasada S., Ferenz K., and Haskell, T. 2002. Conceiving of entities as objects

- and stuff. *Cognition* 83: 141-165.
- Putnam, H. 1975. The meaning of meaning. In *Language, Mind, and Knowledge*, ed. by K. Gunderson. Minneapolis: University of Minnesota Press, 131-193.
- Rey, G. 1983. Concepts and stereotypes. *Cognition* 15: 237-262.
- Rey, G. 2014. Innate and learned: Carey, mad dog nativism, and the poverty of stimuli and analogies (yet again). *Mind and Language* 29: 109-132.
- Rips, L. J., Asmuth, J., and Bloomfield, A. 2013. Can statistical learning bootstrap the integers. *Cognition* 128: 320-330.
- Rips, L. J., Bloomfield, A., and Asmuth, J. 2008. From numerical concepts to concepts of number. *Behavioral and Brain Sciences* 31: 623-642.
- Rips, L. J., and Hespos, S. J. 2011. Rebooting the bootstrap argument: Two puzzles for bootstrap theories of concept development. *Behavioral and Brain Sciences* 34: 145-146.
- Rosch, E., Mervis, C. B., Gray, W., Johnson, D., and Boyes-Braem, P. 1976. Basic objects in natural categories. *Cognitive Psychology* 8: 382-439.
- Shea, N. 2011. Acquiring a new concept is not explicable-by-content. *Behavioral and Brain Sciences* 34: 148-150.
- Smith, C., Carey, S., and Wiser, M. 1985. On differentiation: a case study of the development of size, weight, and density. *Cognition* 21(3): 177-237.
- Smith, E., and Medin, D. 1981. *Categories and Concepts*. Cambridge, MA: MIT Press.
- Spaepen, E., Coppola, M., Spelke, E., Carey, S., and Goldin-Meadow, S. 2011. Number without a language model. *PNAS* 108(8): 3163-8.
- Strevens, M. 2000. The essentialist aspect of naïve theories. *Cognition* 74: 149-175.
- Strevens, M. 2009. Theoretical terms without analytic truths. *Philosophical Studies* 143: 91-100.
- Weiskopf, D. 2008. The origins of concepts. *Philosophical Studies* 140: 359-384.
- Wiser, M. and Carey, S. 1983. When heat and temperature were one. In *Mental Models*, ed. by D. Gentner and A. Stevens. Hillsdale, NJ: Erlbaum, 267-297.

